



ARTICLE



<https://doi.org/10.1057/s41599-022-01181-w>

OPEN

Online public opinion during the first epidemic wave of COVID-19 in China based on Weibo data

Wen-zhong Shi¹, Fanxin Zeng^{1✉}, Anshu Zhang¹, Chengzhuo Tong¹, Xiaoqi Shen², Zhewei Liu¹ & Zhicheng Shi³

As COVID-19 spread around the world, epidemic prevention and control policies have been adopted by many countries. This process has prompted online social platforms to become important channels to enable people to socialize and exchange information. The massive use of social media data mining techniques, to analyze the development online of public opinion during the epidemic, is of great significance in relation to the management of public opinion. This paper presents a study that aims to analyze the developmental course of online public opinion in terms of fine-grained emotions presented during the COVID-19 epidemic in China. It is based on more than 45 million Weibo posts during the period from December 1, 2019 to April 30, 2020. A text emotion extraction method based on a dictionary of emotional ontology has been developed. The results show, for example, that a high emotional effect is observed during holidays, such as New Year. As revealed by Internet users, the outbreak of the COVID-19 epidemic and its rapid spread, over a comparatively short period of time, triggered a sharp rise in the emotion “fear”. This phenomenon was noted especially in Wuhan and the immediate surrounding areas. Over the initial 2 months, although this “fear” gradually declined, it remained significantly higher than the more common level of uncertainty that existed during the epidemic’s initial developmental era. Simultaneously, in the main city clusters, the response to the COVID-19 epidemic in central cities, was stronger than that in neighboring cities, in terms of the above emotion. The topics of Weibo posts, the corresponding emotions, and the analysis conclusions can provide auxiliary reference materials for the monitoring of network public opinion under similar major public events.

¹Otto Poon Charitable Foundation Smart Cities Research Institute and Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hong Kong, China. ²School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu, China. ³Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China. ✉email: fanxin.zeng@connect.polyu.hk

Introduction

Since December 2019, the COVID-19 pandemic has spread rapidly around the world, and many countries have taken measures to interrupt its further spread, such as quarantine and the lockdown of cities (Åslund, 2020; Ren, 2020; Shi et al., 2021; Tiwari et al.; 2021). As a result, to reduce contact opportunities, people have been advised to keep social distancing or stay at home (Tiwari et al., 2021; Welsch et al., 2021). Within this context, communication between people relied mainly on smart devices such as cell phones and computers with the result that social networking platforms such as Weibo became an important platform regarding the enablement of people to socialize and thereby share information. During the period of lockdown measures in China, Weibo became not only a channel enabling people to follow the development of the epidemic and express personal opinions, but also an important tool for Internet users needing such help as medical and health care (Han et al., 2020; Luo et al., 2020). Hence, during the epidemic, the resulting massive social media data, reflected the status of information dissemination on social networking platforms. Such details not only reflected opinions about this major negative public event, but also became vitally important regarding the improvement of emergency response and the subsequent management of public opinions (Liu et al., 2014; Wang and Ye, 2018).

When it comes to public opinion, there is a need to clarify the meaning of the term “public opinion”. This term can be traced back to 1558 when it was first used by Michel de Motaigne. The definition has further evolved with the addition of the highlighting of the word “public”. In the discussion of the phrase: public opinion, the phrase: “public sphere” was used by Jürgen Habermas in his book, he maintained it was the area where “something approaching public opinion can be formed” (Mukerji and Schudson, 1991). According to the American Sociologist Herbert Blumer, public opinion was discussed as a form of collective behavior, made up of those discussing a given public issue at any one time (Habermas et al., 1974). The formation of public opinion begins with the agenda setting by major media outlets all over the world (McCombs and Reynolds, 2002). Nowadays, following the advent of the Internet, the ubiquitous accessibility of social media enables the formation of public opinion both by a broader range of social movements and wider participation (Murphy et al., 2014). Researchers increasingly utilize social media data to both analyze and infer public opinion, by such as identifying topic trends (Murphy et al., 2014; Pearce et al., 2014; Jungherr et al., 2017; Seltzer et al., 2017, 2018; Dubois et al., 2020). Due to the rapid advances in computing technology, the sentiment expressed in the text of social media data attracts the widespread attention of researchers (Cody et al., 2015). In such studies, sentiments tend to be categorized as positive, negative and sometimes neutral (Das and Chen, 2007; Stieglitz and Dang-Xuan, 2013; Schweidel and Moe, 2014; Dhaoui et al., 2017; Ordenes et al., 2017). Based on its use in previous studies, the term “public opinion” is used to describe the collective opinion of society on a certain topic. Hence, more specifically the phrase can also be taken to refer to the sentiment or emotion of the specific components of public opinion.

Based on the social media data witnessed during the epidemic, many scholars have conducted relevant studies from the perspectives of the dissemination of online information, together with changes in residents’ mental states during the pandemic. Several researchers have focused on the development and the transformations of public opinion. Han et al. used both the latent Dirichlet allocation (LDA) and the random forest method to cluster the Weibo text data, together with a study of both the development and transformations of public opinion in several urban agglomerations in China (Han et al., 2020). Zhu et al. used

the LDA method to investigate the temporal and spatial distribution characteristics of related topics during the epidemic in Wuhan (Zhu et al., 2020). However, to study the temporal and spatial evolution of epidemic network public opinion from Wuhan to Hubei Province and even the whole country, Du Yixian et al. constructed a multi-dimensional analysis model, combined with comparative research methods and the Spearman correlation coefficient. The emotion of public opinion was also analyzed (Yixian et al., 2021). Han Keke et al. carried out in-depth information mining using Weibo data during the COVID-19 epidemic situation and by means of a spatial clustering method (Keke et al., 2021), analyzed the differences in network public opinion and key topics in different regions. Gan Yuxiang et al. extracted text feature vectors and context correlation information based on the BERT model and BiGRU model to realize the dynamic modeling of public opinion (Yu-xiang et al., 2021). Tian Yilin et al. conducted in-depth research on the relevant comments on Weibo, constructed the event evolutionary graph of network public opinion during the epidemic, and summarized the corresponding law related to the causal evolution of public opinion (Tian and Li, 2021). Yue Su et al. used Twitter and Sina Weibo data to evaluate the changes in the mental state of residents as expressed through social media, both before and after city closure in Wuhan and Lombardy (Italian city) based on psychological word frequency (Su et al., 2020). Misinformation transmitted by means of the social networks, which had inadvertently caused confusion and eroded public trust during the pandemic, was also explored (Ferrara et al., 2020). Yan Leng et al. used Weibo data and misinformation which was identified by the fact-checking platform of Tencent and thereby found that the evolution of misinformation follows an issue-attention cycle pertaining to topics, such as city lockdown and treatments (Leng, Zhai et al., 2021).

In summary, previous public opinion studies have, more often, only been carried out from the perspective of topic discovery and emotional tendencies, and hence lack the description of fine-grained emotional dimensions. More precisely, specific emotion categories such as happiness, and sadness, have not yet been explored. To fill this particular knowledge gap, this current study uses Weibo data and a fine-grained emotion extraction method based on a dictionary of emotions to study and analyze the development of the urban residents’ public opinion, as witnessed during the first COVID-19 epidemic wave. A specific focus on the perspective of fine-grained emotional dimensions was revealed. Hence, the results can provide both theoretical and technical support, together with auxiliary reference materials, in order to monitor network public opinion under major public events similar to the above, in particular epidemic prevention and control.

Methodology

LDA topic model. LDA is based on the bag-of-words model, hence does not consider the order of words in the document (Thornton et al., 2020). It can be considered that the LDA algorithm process is the inverse process of simulated text generation. In fact, text generation and text topic extraction are two mutually reversible processes. The topic discovery of a text is to find topics that can be described by words or keywords in the text, including the interrelationships between different topics. The text may contain multiple text topics. While the text generation process can be understood as selecting words or keywords that match the topic from the established thesaurus to form the text based on the topic. Therefore, it can be said that text topic discovery and text generation are two opposite processes. The LDA

Table 1 Examples of emotion ontology vocabulary.

Id	Main category	Category	Example
1	happiness (乐)	happy (快乐)	joyful (喜悦), hilarity (欢喜)
2		peace of mind (安心)	relieved (宽心), have a clear conscience (问心无愧)
3	like (好)	respect (尊敬)	respectful (恭敬), respect and love (敬爱)
4		praise (赞扬)	handsome (英俊), superior (优秀)
5		believe (相信)	confide (信任), trust (信赖)
6	anger (怒)	like (喜爱)	admire (倾慕), love (热爱)
7		wish (祝愿)	desire/longing (渴望), bless (保佑)
8		rage (愤怒)	furious (气愤), irritated (恼火)
9	sadness (哀)	sad (悲伤)	down (忧伤), sorrow (悲苦)
10		disappointed (失望)	pity (憾事), despair (绝望)
11	fear (惧)	guilty (疚)	ashamed (惭愧), confess (忏悔)
12		miss (思)	miss (思念), lovesickness (相思)
13		nervous (慌)	panic (慌张), upset (心慌)
14	disgust (恶)	fear (恐惧)	timid (胆怯), afraid (害怕)
15		shy (羞)	shy (害羞), blush (面红耳赤)
16		bored (烦闷)	oppressed (憋闷), fretful (烦躁)
18	surprise (惊)	censure (贬责)	rigid (呆板), vanity (虚荣)
19		jealousy (妒忌)	envious (眼红), jealous (吃醋)
20		suspect (怀疑)	oversensitive (多心), suspicious (生疑)
21		surprised (惊奇)	surprised (惊讶), shocked (大吃一惊)

algorithm is to use the given prior probability parameter value to simulate the inverse process of the text generation process to generate text topics.

LDA model is generally evaluated using the following two indicators: Perplexity and Semantic Consistency (Coherence Score) (Blei et al., 2003; Xie et al., 2018). Perplexity is recognized as a measurement method of information theory. The perplexity calculation of b is defined as related to the entropy of b (b can be a probability distribution or a probability model), which is usually used to compare the effects of probability models. Regarding a model, the smaller the perplexity value, the better the model. The calculation formula for perplexity is as follows:

$$\text{perplexity}(D_{\text{test}}) = \exp \left\{ - \frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\}$$

where M is the number of texts in the test corpus, N_d is the size of the text d , that is, the number of words and $p(w_d)$ represents the probability of the text. According to the bag-of-words model, the probability of a text is the product of the probabilities of all words, and the probability of each word is obtained by the total probability formula (the total probability formula of the topic), so the probability of the text can be obtained.

Although perplexity can describe the uncertainty of topic distribution, it does not show the semantic information of our corpus. The semantic consistency index makes up for this problem. Semantic consistency is another major model for optimal topic number selection. Many studies in China have used this method to determine the number of topics, and topic consistency is the most effective measure of the quality of topics, and it is also one of the important technologies to estimate the number of topics. One of the semantic consistency calculation methods is as follows:

$$C_{\text{UMass}} = \frac{2}{N*(N-1)} \sum_{i=2}^N \sum_{j=1}^{i-1} \log \frac{P(w_i, w_j)}{P(w_j)}$$

where $P(w_i, w_j)$ is the probability of the simultaneous occurrence of the word w_i and word w_j . Therefore, this is the conditional likelihood of a simultaneous occurrence between words. As this simultaneously contains the information of two words, this index can reflect the context, which is better than the perplexity. In previous studies, two indicators were often used to simultaneously

determine the optimal number of topic categories in topic models (Xie et al., 2021). The basic strategy is to minimize the value of perplexity and select the peak of semantic consistency when selecting the optimal number of topic categories. This study has adopted this strategy.

Text emotion extraction. Text emotion extraction has always been an important research issue in the field of natural language processing. Emotion analysis emphasizes the extraction of specific emotion categories, or people’s emotional tendencies toward something such as happiness, anger, sadness, and happiness expressed by users by means of text mining and analysis. In 2013, the 2nd CCF Conference on Natural Language Processing and Chinese Computing considered the complexity of emotions and made requirements for sentiment analysis and evaluation of Chinese Weibo, and divided emotions into seven categories including like (喜好), happiness (高兴), sadness (悲伤), disgust (厌恶), anger (愤怒), fear (恐惧), surprise (惊讶). Based on these seven categories of emotions, the sentiment expressed in the text of Weibo is analyzed.

Bearing in mind the above, the emotion extraction in this current paper uses the emotional ontology vocabulary of the Information Retrieval Laboratory of the Dalian University of Technology. The lexicon divides the human emotions in the text into seven categories. They include “like”, “happiness”, “sadness”, “anger”, “fear”, “disgust”, and “surprise”, each of which basically correspond to the seven types of emotions mentioned above. Examples of emotion ontology vocabulary are shown in Table 1.

When extracting the sentiment characteristics of each Weibo, the cleaned text data is used as input, and the words in the emotion dictionary are used for matching, and the number of keywords in the seven emotion categories of the Weibo is counted to form the Weibo text emotional characteristics. Take the sentence “I scored 100 today, I am very excited and happy” as an example. It can be found that there are 2 words about emotions, namely “excited” and “happy”, and these two words are the subordinate emotional category of “happiness”, so the emotional value of the “happiness” dimension is 2. Suppose that the emotion vector result of a certain Weibo is [‘like’:1, ‘happiness’:1, ‘sadness’:1, ‘anger’:0, ‘fear’:0, ‘disgust’:0, ‘surprise’:0], then the corresponding emotion vector is [1,1,1,0,0,0]. In this way, every Weibo containing emotion can be converted into a one-dimensional

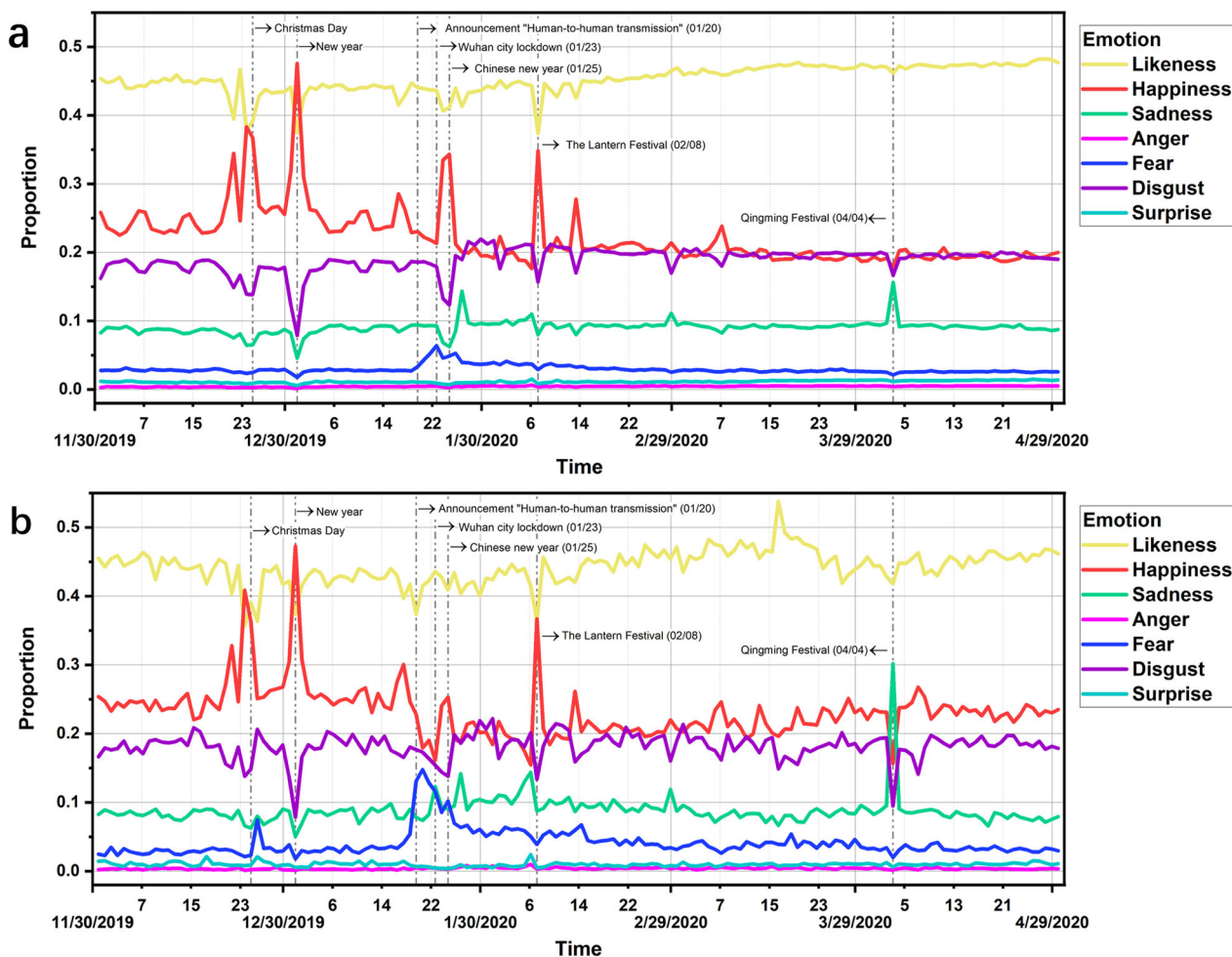


Fig. 1 The emotion distribution over time. The emotion distributions over time during the period from December 2019, to April 2020, **a** across the whole country, **b** in Wuhan city. On each day, the proportion value of a certain emotion is the average value of that emotion in all posts.

emotion vector. Considering that each Weibo blog should have equal weight, the emotion vector can be normalized using the equation below:

$$\tilde{e}_i = \frac{e_i}{\sum_{i=1}^7 e_i}$$

Among them, \tilde{e}_i represents the emotion category i after normalization, and e_i represents the initial emotion value. After normalization, the emotion vector above will be [1/3,1/3,1/3,0,0,0].

Results

Daily emotional changes of Weibo users. In order to daily observe the emotional changes of Weibo users, the average operation of the city and the date is taken. Then, the daily change of seven emotion categories in each city is obtained. Figure 1 demonstrates the emotion distribution over time, both over the whole country and in Wuhan during the epidemic period.

As shown in Fig. 1, The distributions of emotions over time both across the country and in Wuhan were similar. All the emotional categories were generally stable over the period studied. At certain time points, certain emotions presented large fluctuations. Specific peak periods include December 22, 2019 to December 25, 2019; January 1, 2020; January 18, 2020 to January 24, 2020; February 8, 2020; and April 4, 2020. The Christmas holiday periods which ran from December 22, 2019, to December 25, 2019, and January 1, 2020, being New Year’s Day, can explain the significant increase in

the proportions of both the two positive emotions categories, “like” and “happiness”. Another notable period is from January 18, 2020 to January 24, 2020. On January 20, 2020, Zhong Nanshan announced that “there must be human-to-human transmission”. In the following days, the pandemic gained great attention. During this period, the emotions in Wuhan and China fluctuated greatly. Specifically, in the case of emotion “fear”, Wuhan experienced a rapid climb and peaked on January 21 before gradually declining. Unlike in Wuhan city, the emotion “fear” in the whole country did not increase on a large scale, and the growth trend was relatively late. Both the overall and the peak appeared nearly two days later than the appearance in Wuhan city. Both peaks were followed by a rise in positive emotions, especially the emotion “happiness”. It is noted that the peak of the emotion “happiness” occurred on January 25, the day of the Chinese New Year. A similar situation occurred on February 8, which was the Lanterns Festival, a Chinese festival. On that day, both in Wuhan and across the country, the emotion “happiness” increased dramatically. Another notable change is that on April 4, 2020, there was a surge in the emotion “sadness” both across China and in Wuhan. This particular day was Tomb-sweeping Day, also known as Ching Ming Festival, another traditional Chinese festival. During Qingming, Chinese families visit the tombs of their ancestors to clean the gravesites, pray to their ancestors and make ritual offerings. On April 4, 2020, China’s State Council has held national mourning activities to express the deep sorrow for those who lost their lives in the fight against COVID-19 and those who died from COVID-19.

The spatial and temporal difference in terms of emotion “fear”.

As can be seen, in the early stage of the COVID-19 pandemic, the emotion “fear” increased across the whole country, but more especially in the city of Wuhan, in which a surge was experienced. To better understand the spatial difference in emotion distributions, several major cities in China have been selected. Included were Guangzhou, Shenzhen, Shanghai, Beijing, Chengdu, all of which, are not simply the centers of city clusters in China, but are also the major destination cities of the Wuhan emigration population during half of the month prior to the Wuhan lock-down. The neighboring cities of these central cities were used to compare the spatial difference in emotion distributions. Considering that Guangzhou and Shenzhen are very geographically close, the same neighboring cities were selected for further comparison. In order to eliminate the influence of the number of Weibo posts on the emotion extraction results, the Weibo data of neighboring cities were merged. Based on the emotion vector results extracted from the Weibo texts, the “fear” distributions both within these central cities and their neighboring cities were calculated (see Fig. 2).

It is seen from Fig. 2 that all cities had witnessed both a surge and peak of “fear” during the period from January 20, 2020, to January 27, 2020. The peak values were 4–5 times the usual level, especially in Wuhan, in which the ratio was more than 6 times. Specifically, the peaks appeared on January 20, 2020, and January 24, which were, respectively, the days following the announcement of “human-to-human transmission” and the lock-down of

Wuhan city. In the cases of peak times, the neighboring cities were affected later than the central cities. Subsequently, “fear” in all cities experienced a downward trend, however, the overall level remained at 1–2 times the pre-epidemic level. Even after Wuhan was reopened on April 8, a major victory in China’s fight against the epidemic, the “fear” level in all cities was still slightly higher than the pre-epidemic level. Moreover, during the period from January 20, 2020 to March 1, 2020 when confirmed cases continued to increase every day, the central cities presented a higher proportion of the emotion “fear” than the neighboring cities, hence appearing to indicate that citizens in these specific central cities reacted to the COVID-19 more strongly than those in the neighboring cities, in terms of “fear”. Even in the Pearl River Delta region, which has a more balanced development among the cities, Guangzhou and Shenzhen still presented a higher “fear” value than the neighboring cities. Hence, the difference between central cities and the smaller neighboring cities was confirmed, and public opinion management should be thus further encouraged to focus on those cities playing a dominant role in the city clusters of the regions.

The impact on the emotions of residents in the face of the residential city closure policy.

In order to explore the impact of the city closure policy on the overall emotion of residents in Wuhan, the emotional distributions of Weibo posts related to “lockdown (封城)” in Wuhan both before and after the policy release on January 23, 2020, were calculated (see Fig. 3).

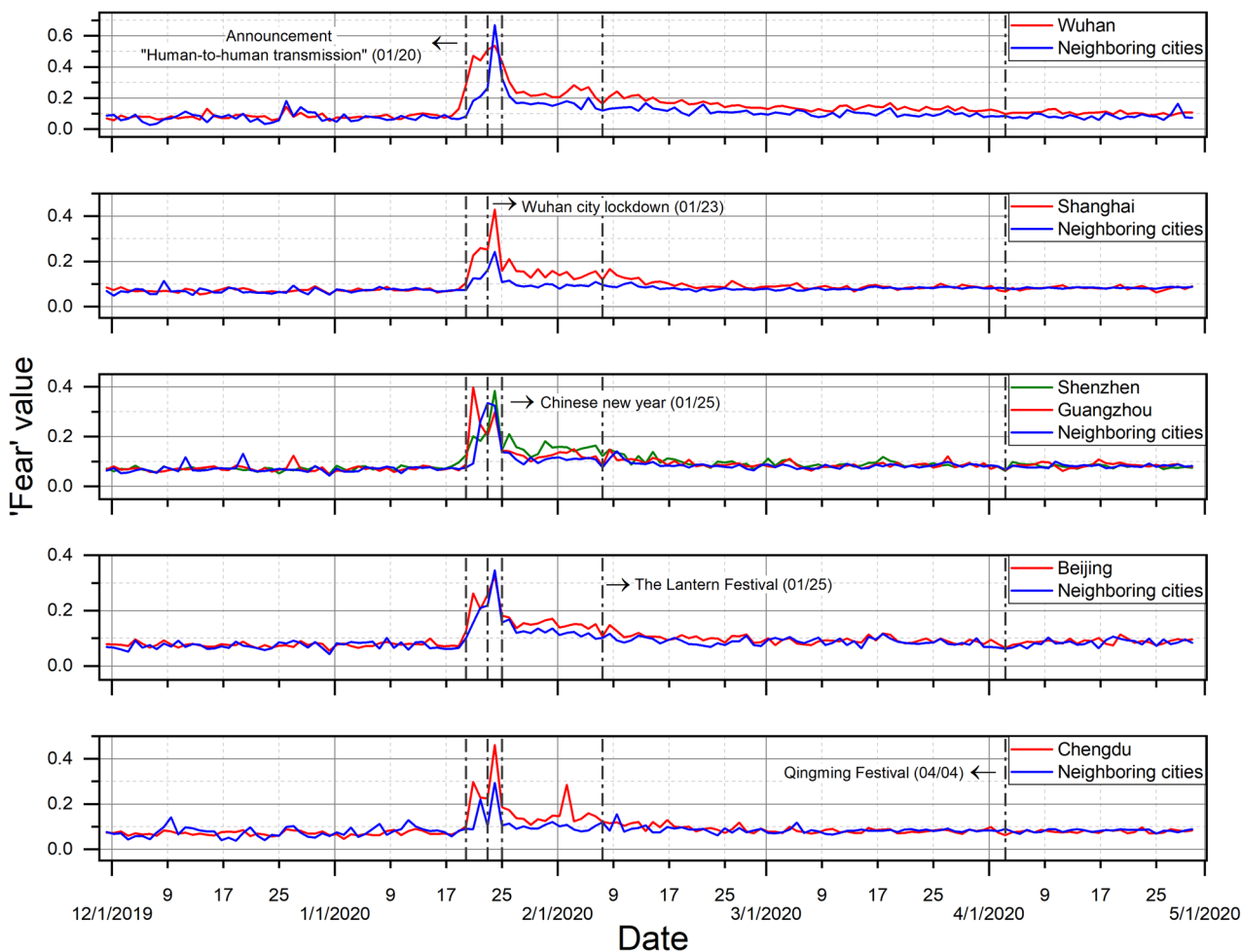


Fig. 2 The distribution of the emotion “fear” over time in different cities. This line chart compares the changes in the emotion “fear” of different central cities and their neighboring cities.

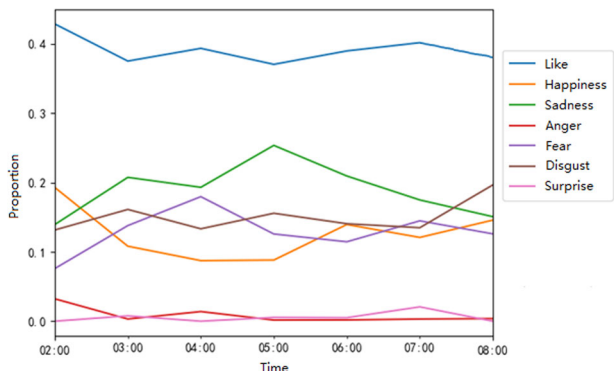


Fig. 3 Emotional distribution of Weibo posts related to 'Lockdown'. The emotional distribution of Weibo posts related to 'Lockdown' during hours after the release of the lockdown policy on January 23, 2020.

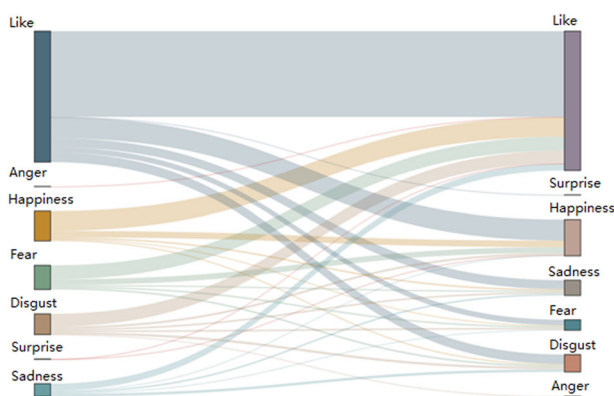


Fig. 4 Changes of emotions of 1149 users before and after lockdown. The changes of emotions of 1149 users before and after lockdown, with the widths of flows representing the numbers of users.

During two hours after the release of the policy, public sentiment towards the lockdown policy was relatively negative. Specifically, “fear” and “sadness” both significantly increased, suggesting that on the whole, residents were uneasy about the sudden lockdown arrangement in the short term.

To explore the impact of Wuhan’s lockdown policy on personal emotional changes, the period from 00:00 on January 20, 2020, to 24:00 on January 26, 2020, was re-selected as the research time range, in which Wuhan issued the relevant lockdown decision notice at 3:00 a.m. on January 23, 2020. Hence, by that notice time, the time range can be divided into two time periods: ‘before’ and ‘after’ the lockdown. First, 1449 people who posted emotional tweets before and after the lockdown were selected. Before and after lockdown, the average values of the emotional distribution of each user’s multiple posts were used as the emotional distribution of the user, and the emotion with the highest proportion in the emotional distribution of each user was further selected as the dominant emotion of the Weibo user. Accordingly, the result is found to be that each user’s mood tended to change from one dominant emotion category before the lockdown to another specific dominant emotion category after the lockdown.

As shown in Fig. 4, the emotion of quite a few users changed greatly before and after the lockdown. The two emotions, “like” and “happiness”, were still regarded as positive emotions. The overall number of users obtaining positive emotions increased, despite some users who had positive emotions before the lockdown changing to negative emotions. Correspondingly, the

number of people holding negative emotions decreased. In specific emotional categories, “fear”, “disgust”, “anger”, “sadness” and “surprise”, it was apparent that a certain number of users switched to the positive category. Moreover, among these 1149 users, the number of people who had felt fear decreased. Given that January 25 is the Chinese New Year, the improvement in sentiment in the analysis may be influenced both by residents’ confidence in the governmental policy and also the festival atmosphere of the Chinese New Year.

Combined with topic modeling results. The LDA topic model was used to classify the Wuhan city Weibo posts on January 23, 2020. Figure 5 shows both the variation of perplexity and also semantic consistency under the different numbers of topic categories.

As can be seen, when the number of topic categories is between 4 and 10, the perplexity tends to fluctuate and stabilize. However, with an increase in the number of topics, the perplexity of the model, likewise, gradually increases, indicating that the greater the number of post categories, the greater the uncertainty of the model results. The semantic consistency fluctuated with a range of [0.465,0.505] and reached the peak value when the number of topic categories reached 6, 9, 13 and 15, respectively.

The keywords in Table 2 are the results of the LDA topic model, and the topic name is determined by the semantic information of the keywords. It is seen that the LDA topic model is able to extract potential topics in Weibo discussions. In order to facilitate further observation of the topic classification results, a corresponding word cloud is drawn using the classified Weibo text data (see Fig. 6).

It is seen, that basically, the content of the Weibo posts of each category is consistent with the distribution of the keywords. In category 0, the topic is summarized as “traffic measures”, and the word cloud rendering shows that the main content was traffic operation suspension measures taken by Wuhan on the day of the lockdown. On that day, Wuhan had temporarily closed the airport, railway station, and long-distance passenger terminal, and also ordered buses, subways, and ferries to suspend operations. The topic of category 1 is summarized and headed as “Chinese New Year”. The main content of Weibo posts includes key phrases such as “New Year” and “going home”. Of interest is that the most popular word in the word cloud is “hope”. A search of all the Weibo posts reveals that many sentences contain “hope” in the text, such as “Hope the epidemic would end as soon as possible”, while some users expressed regret that they “could go home early”. The topic summarized in category 2, is the “epidemic” itself, and it is clearly and visually emphasized by the word cloud. Words signaling the development of COVID-19, are clearly emphasized by the use of keywords such as “confirmed”, “nationwide”, and “hospital”. In category 3, the major topic is “life”, emphasized and illustrated by the use of such keywords in the word cloud visualization results, including “life”, “citizen”, “peace” and “hope”. In addition, key phrases such as “hold on” and “stick to”, illustrate positive emotions. In category 4, the topic is summarized as “materials”. The discussion posts in this category focus mainly on keywords such as “masks”, “supermarkets”, “supplies” and “going out”. A certain number of words or phrases such as “cannot buy” also appeared. According to news reports at that time, there was a buying frenzy at major supermarkets on the day of the lockdown. In category 5, the topic is summarized as “social relations”. As can be seen, “city”, “friends”, “country”, “government” are the main keywords in the above category. In this closed city, given the above, social relationship also becomes an important concern.

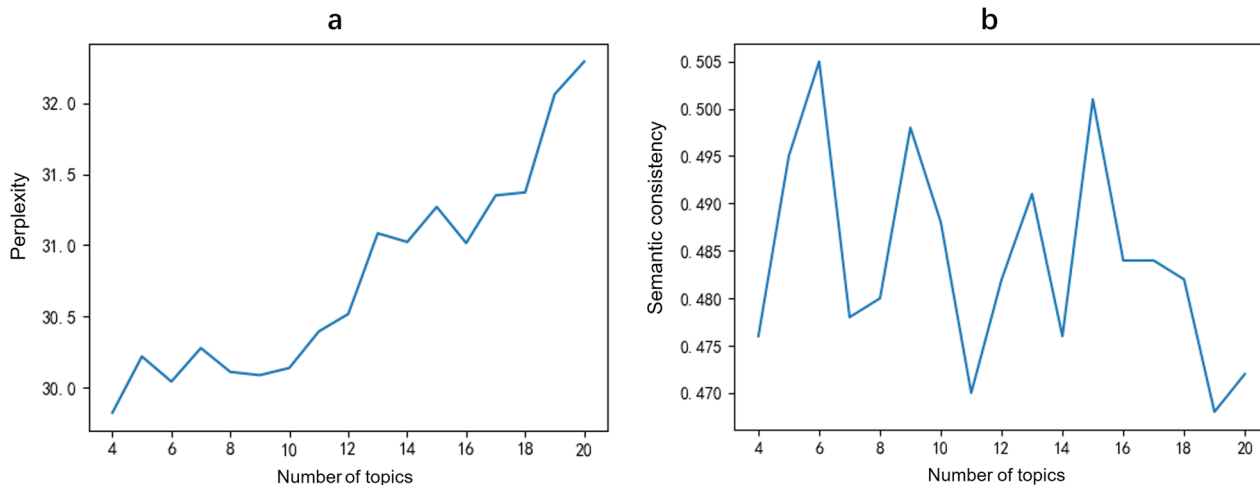


Fig. 5 Evaluations of LDA topic model. The evaluation indicators of the LDA topic model. **a** Perplexity and **b** semantic consistency vary with the number of topics.

Table 2 Topics-keywords distribution on January 23 in Wuhan.

No.	Topics	Keywords
0	Traffic measures (交通措施)	lockdown (封城), subway (地铁), bus (公交), 23, suspend (暂停), operation (运营), God bless (天佑), cheer up (加油)
1	Chinese new year (中国新年)	lockdown (封城), hope (希望), New Year (过年), go home (回家), go to work (上班), go back (回去), at home (家里), family (家人)
2	Epidemic (疫情)	COVID (新冠肺炎), nationwide (全国), hospital (医院), confirmed (确诊), medical case of illness (病例), doctor (医生), pandemic (疫情), virus (病毒)
3	Life (生活)	cheer up (加油), life (生活), peace (平安), citizen (市民), hold on (挺住), protect (保护), health (健康), stand fast (坚守)
4	Materials (物资)	mask (口罩), supermarket (超市), at home (在家), materials (物资), protection (防护), disinfect (消毒), alcohol (酒精), cannot buy (买不到)
5	Social relationship (社会关系)	city (城市), friend (朋友), hope (希望), nation (国家), government (政府), family (家人), care about(关心), get ill (生病)

To further explore the distribution of emotions category by category, this paper presents statistics regarding emotions of Weibo posts based on topics in the classification results. Figure 7 shows the distribution of emotions based on each topic category. The distributions of these emotions illustrated similarities and also vast differences. Firstly, the emotions “anger” and “surprise” in all the topic categories are similar, with the proportions very small. Hence the focus of this discussion will not be on the above two emotions. The main topic categories 0, 3, and 5 are proportionally higher in “like” and include “traffic”, “life”, and “social relationship”. According to the topic-keywords distribution results, the keywords found in these categories include the words and phrases: “God bless”, “come on”, “friends” and “concern”, each obviously expressing positive emotions. However, the high frequency of “Peace” makes category 3 higher in the proportion of “happiness” emotion. Category 2 illustrates that users were highly concerned about the developmental rate of the COVID-19 pandemic; hence, category 2 has the highest proportion of the emotion “fear”. Categories 1 and 4, indicate the proportion of “fear” to be slightly higher, with the topics being “New year”, “going home” and “material” related to personal issues. In category 4, the proportion of the emotion “disgust” is also high, which reflects the residents’ attention to “material”, as well as the “fear” and “disgust” emotions possibly caused by the insufficient reserves of “material” for life in the epidemic area under lockdown. Category 5, presents the emotion “disgust” in a high proportion. Keywords such as “government”, “country” and “care about” were used. A search of topics related to the above,

and as would be expected, reveal both negative and positive posts, relating to the local government. For instance, mentioned was “cooperating with the local government’s arrangements”, while others also “scolded”, “cursed” and “criticized” what they saw as the local government’s arrangements. Hence it was apparent that on the day of the lockdown, residents in Wuhan, as would be possibly expected in any diverse society, were relatively concerned about the local government’s work arrangements in the area’s time of need.

Discussion

In a major national public health event such as the COVID-19 pandemic, information dissemination is heavily dependent on social media platforms such as Weibo. However, the massive amount of social media data achieved provides a valuable opportunity from which public opinion can be gathered. The monitoring of public opinion, to a reasonable degree, enables some understanding of the concerns and needs endured by residents under such special circumstances. Such knowledge contributes to better enabling the relevant authorities to take corresponding response measures as regards the planning of necessary features which, as far as possible, will better enable work and normal domestic functions. The research information presented in this paper aims, by means of Weibo data, to present an analysis of both the course and development of online public opinion in China, during the period from December 2019, to April 2020. A text emotion extraction method based on the

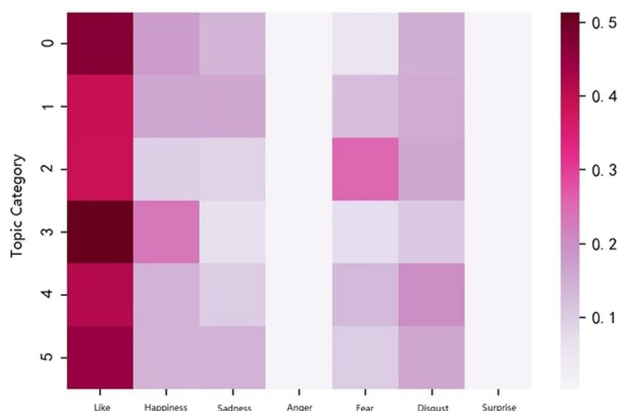


Fig. 7 Emotion distributions in different topic categories. The comparison of emotion distributions by topic category.

development of the epidemic and the specific measures to restrict the movement of people, but also the supply of household goods during the closure. Therefore, it is believed that in a similar situation to closed management, the assurance of the supplies is likely to be effective in relieving the tension and fear of the residents. Additionally, the Chinese New Year festival was a hot topic due to its proximity. Surprisingly, the results showed that social relationships were also an important topic. It became apparent that in a tense atmosphere, people tended to look for support and help from links with both others and the government. Combined with the corresponding emotion extraction results, it can be conjectured that the local government should increase policy publicity and ensure efficient information sharing when making relevant work arrangements.

This study provides a more refined perspective for public opinion analysis, but future works need also be enabled to focus on the improvements of the text emotion extraction method. The emotional lexicon utilized in this paper focuses on adjectives in the text, however, as it is of note that the textual emotional expression does not always depend on the use of adjectives. A more comprehensive vocabulary of sentiments or a vocabulary of tweet-specific expressions would make the results more convincing. A more fine-grained approach to short text topics can also be further explored.

Data availability

This research was based on the dataset: Weibo-COV: A Large-Scale COVID-19 Social Media Dataset from Weibo (Hu et al., 2020). More information can be found in the Github project: <https://github.com/nghuyong/weibo-public-opinion-datasets>.

Received: 4 December 2021; Accepted: 25 April 2022;
Published online: 06 May 2022

References

Åslund A (2020) Responses to the COVID-19 crisis in Russia, Ukraine, and Belarus. *Eurasian Geogr Econ* 61(4–5):532–545

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3(4–5):993–1022

Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM (2015) Climate change sentiment on Twitter: an unsolicited public opinion poll. *PLoS ONE* 10(8):e0136092

Das SR, Chen MY (2007) Yahoo! for Amazon: sentiment extraction from small talk on the Web. *Manag Sci* 53(9):1375–1388

Dhaoui C, Webster CM, Tan LP (2017) Social media sentiment analysis: lexicon versus machine learning. *J Consum Mark* 34(6):480–488

Dubois E, Gruz A, Jacobson J (2020) Journalists' use of social media to infer public opinion: the citizens' perspective. *Soc Sci Comput Rev* 38(1):57–74

Ferrara E, Cresci S, Luceri L (2020) Misinformation, manipulation, and abuse on social media in the era of COVID-19. *J Comput Soc Sci* 3(2):271–277

Habermas J, Lennox S, Lennox F (1974) The public sphere: an Encyclopedia Article (1964). *New German Critique* 3:49

Han X, Wang J, Zhang M, Wang X (2020) Using social media to mine and analyze public opinion related to COVID-19 in China. *Int J Environ Res Public Health* 17(8):2788

Hu Y, Huang H, Chen A, Mao X-L (2020). Weibo-COV: a large-scale COVID-19 Social Media dataset from Weibo. arXiv preprint.

Jungherr A, Schoen H, Posegga O, Jürgens P (2017) Digital trace data in the study of public opinion. *Soc Sci Comput Rev* 35(3):336–356

Keke H, Ziyao X, Zhe L, Junming L, Xiaodong Z (2021) Research on public opinion analysis methods in major public health events: take COVID-19 epidemic as an example. *J Geo-inf Sci* 23(2):331–340

Liu Q, Gao Y, Chen Y (2014) Study on disaster information management system compatible with VGI and crowdsourcing. 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA), IEEE

Leng Y, Zhai Y, Sun S, Wu Y, Selzer J, Strover S, Zhang H, Chen A, Ding Y (2021) Misinformation during the COVID-19 outbreak in China: cultural, social and political entanglements. *IEEE Trans Big Data* 7(1):69–80

Luo C, Li Y, Chen A, Tang Y (2020) What triggers online help-seeking retransmission during the COVID-19 period? Empirical evidence from Chinese social media. *PLoS ONE* 15(11):e0241465

McCombs M, Reynolds A (2002) News influence on our pictures of the world. Media effects: advances in theory and research, 2nd edn. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, USA, pp. 1–18

Mukerji C, Schudson M (1991) Rethinking popular culture: contemporary perspectives in cultural studies. University of California Press.

Murphy J, Link MW, Childs JH, Tesfaye CL, Dean E, Stern M, Pasek J, Cohen J, Callegaro M, Harwood P (2014) Social Media in Public Opinion Research Executive Summary of the Aapor Task Force on emerging technologies in public opinion research. *Public Opin Q* 78(4):788–794

Ordenes FV, Ludwig S, De Ruyter K, Grewal D, Wetzels M (2017) Unveiling what is written in the stars: analyzing explicit, implicit and discourse patterns of sentiment in Social Media. *J Consum Res* 43(6):875–894

Pearce W, Holmberg K, Hellsten I, Nerlich B (2014) Climate change on Twitter: topics, communities and conversations about the 2013 IPCC Working Group 1 Report. *PLoS ONE* 9(4):e94785

Ren X (2020) Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States. *Eurasian Geogr Econ* 61(4–5):423–434

Schweidel DA, Moe WW (2014) Listening in on Social Media: a joint model of sentiment and venue format choice. *J Mark Res* 51(4):387–402

Seltzer EK, Horst-Martz E, Lu M, Merchant RM (2017) Public sentiment and discourse about Zika virus on Instagram. *Public Health* 150:170–175

Shi W, Tong C, Zhang A, Wang B, Shi Z, Yao Y, Jia P (2021) An extended Weight Kernel Density Estimation model forecasts COVID-19 onset risk and identifies spatiotemporal variations of lockdown effects in China. *Commun Biol* 4:1

Stieglitz S, Dang-Xuan L (2013) Emotions and information diffusion in Social Media—sentiment of microblogs and sharing behavior. *J Manag Inf Syst* 29(4):217–248

Stieglitz S, Mirbabaie M, Ross B, Neuberger C (2018) Social media analytics—challenges in topic discovery, data collection, and data preparation. *Int J Inf Manag* 39:156–168

Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, Liu T, Gong W, Zhu T (2020) Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter. *Int J Environ Res Public Health* 17(12):4552

Thornton A, Meiners B, Poole D (2020) Latent Dirichlet Allocation (LDA) for anomaly detection in avionics networks. 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), IEEE

Tian Y, Li X (2021) Analysis on the evolution path of COVID-19 network public opinion based on the event evolutionary graph. *Inf Stud: Theory Appl* 44(3):76–83

Tiwari A, Sandalwala J, Joshi M, Gupta AK (2021) COVID-19: government intervention and post covid complications in India. *J Pharm Res Int* 33(37b):168–178

Wang ZY, Ye XY (2018) Social media analytics for natural disaster management. *Int J Geogr Inf Sci* 32(1):49–72

Welsch R, Wessels M, Bernhard C, Thönes S, Von Castell C (2021) Physical distancing and the perception of interpersonal distance in the COVID-19 crisis. *Sci Rep* 11:1

Xie R, Chu SKW, Chiu DKW, Wang Y (2021) Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis. *Data Inf Manag* 5(1):86–99

- Xie T, Qin P, Zhu L (2018) Study on the topic mining and dynamic visualization in view of LDA model. *Mod Appl Sci* 13(1):204
- Yixian D, Jiapeng X, Linying Z, Yingxu H, Jie S (2021) Analysis and visualization of multi-dimensional characteristics of network public opinion situation and sentiment: taking COVID-19 epidemic as an example. *J Geo-inf Sci* 23(2):318–330
- Yu-xiang G, Ya-bo W, Jun-xiao X, Ruo-qi Z, Shu-ning X, Yi-bo G (2021) Public opinion evolution analysis of “COVID-19 epidemic” based on sentiment feature. *J Graph* 42(2):222–229
- Zhu B, Zheng X, Liu H, Li J, Wang P (2020) Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. *Chaos Solitons Fractals* 140:110123

Acknowledgements

This study was supported by the National Key R&D Program of China (2019YFB2103102), Research Grant Council, HKSAR Government (C5079-21G), Innovation and Technology Fund, HKSAR Government (ITP/041/21LP), and Otto Poon Charitable Foundation Smart Cities Research Institute, The Hong Kong Polytechnic University (Work Program: CD03).

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical assessment is not required prior to conducting the research reported in this paper, as the present study does not have experiments on human subjects and animals, and does not contain any sensitive and private information.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Fanxin Zeng.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022